DOCUMENT RESUME

ED 089 098 CE 001 153

| | |
|---|---|
| AUTHOR | Denton, William T. |
| TITLE | Student Evaluation in Vocational and Technical Education. |
| INSTITUTION | Ohio State Univ., Columbus. Center for Vocational and Technical Education. |
| SPONS AGENCY | National Inst. of Education (DHEW), Washington, D.C. |
| REPORT NO | IS-97 |
| PUB DATE | 73 |
| NOTE | 34p. |
| EDRS PRICE | MF-$0.75 HC-$1.85 PLUS POSTAGE |
| DESCRIPTORS | Achievement Tests; Aptitude Tests; Criterion Referenced Tests; Interest Tests; Literature Reviews; Student Attitudes; *Student Evaluation; *Technical Education; *Test Construction; Testing Programs; *Test Selection; *Vocational Education |

ABSTRACT

The challenge of student evaluation in the area of vocational and technical education is viewed by the author as being highly related to the problem of measurement. The purpose of this paper is to review some of the instruments currently used to measure student attainments and/or attitudes. Following a discussion of general aspects (amount of testing, level of test difficulty, and test administration) which should be considered before embarking on an evaluation program, the author provides a review of some commonly used measures, divided into achievement and aptitude tests, and interest tests. For situations for which one of the available measures is not appropriate, it may be necessary to develop a test locally. Detailed guidelines are offered to encourage this endeavor in the second half of the document. For companion documents covering evaluation of facilities, program, and personnel in vocational and technical education see, CE 000 990, CE 000 988, and CE 001 133. (Author/SA)

FEB - 1974

Information
Series No. 97

ED 089098

# STUDENT EVALUATION

IN

## VOCATIONAL AND TECHNICAL EDUCATION

ERIC

Clearinghouse on Vocational
and Technical Education

CE001153

Information Series No. 97

# STUDENT EVALUATION

## IN

## VOCATIONAL AND TECHNICAL EDUCATION

William T. Denton
Dallas Independent School District
Dallas, Texas

ERIC Clearinghouse on Vocational and Technical Education
The Center for Vocational and Technical Education
The Ohio State University
1960 Kenny Road Columbus, Ohio 43210

1973

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402

The material in this publication was prepared pursuant to a contract with the National Institute of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to the American Vocational Association for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either the American Vocational Association or the National Institute of Education.

## CONTENTS


INTRODUCTION. . . . . 1

GENERAL CONSIDERATIONS. . . . . 2
Amount of Testing. . . . . 2
Level of Test Difficulty . . . . . 2
Administration of the Test . . . . . 3

WIDELY DISSEMINATED MEASURES. . . . . 3
Achievement and Aptitude Tests . . . . . 4
Interest Tests . . . . . 10
Summary. . . . . 12

LOCALLY DEVELOPED MEASURES. . . . . 12

EPILOGUE. . . . . 24

BIBLIOGRAPHY. . . . . 27

## FOREWORD

Vocational and technical education has enjoyed high visibility during the past few years and with it increased pressure to account for expenditures and to justify programs. As a result, educators are ever alert for effective means of evaluating their educational programs. This publication and its three companion documents (Program Evaluation in Vocational and Technical Education, Personnel Evaluation in Vocational and Technical Education, and Facilities Evaluation in Vocational and Technical Education) provide educational practitioners with a review and synthesis of the most important works in evaluation as it applies to vocational and technical education.

In Student Evaluation in Vocational and Technical Education, the author looks at some general considerations related to testing; discusses the most widely disseminated achievement, aptitude, and interest tests; and offers guidelines for those interested in developing their own tests.

The profession is indebted to William T. Denton for his scholarship in the preparation of this report. Recognition is also due Gordon Law, Department of Urban Education, Rutgers--the State University; and Donald L. Rathbun, Associate Director, American Vocational Association for their critical review of the manuscript prior to final revision and publication. Paul E. Schroeder coordinated the publication's development, and Alice J. Brown and Paula Kurth provided the technical editing.

Robert E. Taylor
Director
The Center for Vocational and Technical Education
ERIC Clearinghouse on Vocational and Technical Education

# INTRODUCTION

The challenge of student evaluation in the area of vocational and technical education is viewed by the author as being highly related to the problem of measurement. The purpose of this paper, therefore, is to review some of the instruments currently used to measure student attainments and/or attitudes.

The intended audience is the evaluator who finds himself in the position of either developing or finding measures of achievement for vocational and technical education students. The paper provides a quick review of some commonly used measures, how to develop them locally, and a source of further information for the evaluator.

For organizational purposes, the field of measurement is divided into two major categories: (1) standard, widely disseminated measures; and (2) locally developed measures. The first category is characterized as being appropriate to a more generalized situation while the second category is characterized as being more situation specific. The field of criterion-referenced measures has been placed in the second category.

There continues to be a great deal of interest in student testing. Older tests are being revised and new tests are constantly being developed and piloted. For an excellent source of information about various tests, the reader is referred to (Buros, 1972). The tests discussed in this paper are by no means the only ones available, however, in most cases they are the more commonly used ones.

The first section of the paper discusses some general considerations that every evaluator should heed before entering into any testing program. The points brought out will contribute considerably toward the collection of information useful in evaluating students.

The second section discusses widely disseminated measures commonly used to evaluate students. A brief discussion of the most often used instruments is provided including a summary of some reviews. The instruments included are classified as achievement (including aptitude) and interest measures.

The third section discusses the challenges inherent in attempting to develop measures locally. The discussion starts with the development of behavioral objectives and leads into how to write and test criterion measures of the behavioral objectives.

## GENERAL CONSIDERATIONS

With the increased emphasis on evaluation of students has come an increased use of tests. Carpenter and Rapp (1972:1) admonishes evaluators that, "Because of the importance being placed on test results, there is an urgent need to observe good testing practices." They point out that considerations for good testing, regardless of the specific test used, are relative to: (1) amount of testing, (2) level of test difficulty, and (3) administration of the test.

### Amount of Testing

Students in individualized programs will be tested a great deal, especially if there are criterion-referenced measures to determine mastery, all of which reduce the amount of time available for instruction. No pat answer is known to the question of optimal testing; however, the evaluator should be aware of the problem and minimize testing time while still obtaining the desired amount of information.

### Level of Test Difficulty

Evaluators are aware of the difficulties encountered when a test too difficult for the student population is administered: many students score at or below chance level. One possible solution might be to give a test requiring a lower level of ability. However, this presents at least two problems, especially at the upper grades. First is the problem of how to interpret the scores, and second is the chance of testing with low interest materials. The second problem may affect student motivation and consequently, performance.

Administration of the Test

It is wise for the testing program to be planned well in advance so that the building principal can be advised of the extent of the disruption of normally scheduled activities. The evaluator must make sure that those taking the test, as well as those giving the test, have all the necessary supplies, which include pencils and timing devices. Prior to the testing, each tester should be given a copy of the test and written instructions outlining what is required of him. If the testing procedures are out of the ordinary, the tester should practice giving the directions. Testing facilities should be arranged so that those taking the test will not be uncomfortable nor distracted by answers other students might give. It is routine to have a practice session precede the test to help those taking the test understand better what it is they are expected to do. It is extremely important that all of the necessary identification information be properly included on each test. The actual testing should be monitored to identify and hopefully eliminate any gross discrepancies in the testing procedure. Finally, a random sample of tests should be rescored to determine if the scoring error rate is excessive (Carpenter and Rapp, 1972).

## WIDELY DISSEMINATED MEASURES

For discussion purposes, this section will be divided into two broad types of student measures: (1) achievement and aptitude tests, and (2) interest tests. These two categories are not intended to be exhaustive, but rather to include those measures most commonly used to measure student achievement and interest in the field of vocational and technical education.

Seibel (1968), in discussing achievement, scholastic aptitude, and intelligence tests, points out that the three categories actually have more similarities than differences. They all measure learned skills and abilities, they all can be used to predict learning, and they all measure "intelligent behavior."

Seibel warns that

> Unless the teacher or test user looks beneath the superficial descriptions of a given test and actually examines the tasks to be performed, there is no assurance that the test results will be meaningful or useful for him. In fact, the results may present a completely false picture of student or class accomplishments (1968:265).

With this advice in mind, the next step is to survey some selected achievement tests related to vocational and technical education.

## Achievement and Aptitude Tests

Although the Illinois Battery developed under the guidance of Baldwin (1970) has been focused on the post-secondary level of vocational and technical education, it is probably relevant to other levels. A great deal of effort has been extended in developing instruments to measure achievement in vocational and technical education. The work was completed under a four-year grant from the U.S. Office of Education (USOE). The project was a joint venture of the University of Illinois and North Carolina State University.

The following tests have been produced (Baldwin, 1970:2):

1) Achievement Test for Machinist

2) Auditory Achievement Test for Machinist

3) Achievement Test for Radio and Television Servicing

4) Visual Diagnostic Test for Television Servicing

5) Achievement Test for Air Conditioning, Heating and Refrigeration

6) Achievement Test for Automotive Mechanics

7) Auditory Achievement Test for Automotive Mechanics

8) Achievement Test for Electrical Installation and Maintenance

9) Achievement Test for Data Processing Technology--Business

10) Achievement Test for Data Processing Technology--Scientific

11) Achievement Test for Electronics Technology

The selected curricula, from which the achievement tests were developed, was already field-tested by the Curriculum Laboratory of the North Carolina Department of Community Colleges. Members of the item-writing pool were asked to estimate the percentage of time devoted to each of the subdivisions of the curriculum. From this consensus estimate, the proportion of items necessary for each subdivision of each curriculum was determined. The items were then written by the various members of the committee and discussed by the committee as a whole. Only those items mutually agreed upon were included in the first version of the test. With the exception of the air conditioning, heating and refrigeration area, two paper-and-pencil tests were made for each curriculum. The tests were given to a sample of students and the results analyzed. From the information gained during the analysis, the tests were revised. For a technical discussion of the analyses the reader is referred to The Development of Achievement Measures for Trade and Technical Education (Baldwin, 1970).

The Ohio Trade and Industrial Education Services (1972) have developed a Trade and Industrial Education Achievement Test program which consists of the California Short Form Test of Academic Aptitude plus the following Trade Achievement Tests:

1) Auto Body,

2) Automotive Mechanics,

3) Basic Electricity,

4) Basic Electronics,

5) Carpentry,

6) Cosmetology,

7) Dental Assisting,

8) Machine Trades,

9) Mechanical Drafting,

10) Printing,

11) Sheet Metal, and

12) Welding.

Each of the Trade Achievement Tests is developed from a course outline. The course outline is developed from an analysis of the trade and inputs obtained from a committee appointed by the assistant director of trade and industrial education. The committee consists of a representative of the state supervisory staff, a teacher-educator, a local supervisor of trade and industrial education, a selected teacher of the course, and a representative of the Ohio Trade and Industrial Education Services, Instructional Materials Laboratory. The committee then uses the course outline to develop test items. Items are reviewed and those upon which there is agreement are used for the initial testing. After each testing period, revisions are made where necessary. For a technical discussion of the analyses of the tests, the reader is referred to Using the Results of the Trade and Industrial Education Achievement Test Program (Ohio Trade and Industrial Education Services, 1972).

Any of the tests included in the two sets previously discussed are likely to be of value to the evaluator. There are two possible limitations of which the evaluator should be cognizant. First, as pointed out earlier, are the tasks to be performed relevant to the local situation? Since both sets of tests were developed from either existing curricula or developed curriculum guides, a local evaluator should, by getting the relevant materials, be able to ascertain whether or not the tests are applicable to the local situation. Second, are the referent groups used to "norm" the tests similar enough to make valid comparisons to the local student population (Seibel, 1968)? In the opinion of this author, the norms developed for either set of these tests should not be interpreted as representative of a national population. It would seem appropriate, especially for those using the tests who are not from the states included in the

norming sample, to consider developing a set of local norms. As Angoff points out, "...what constitutes satisfactory performance, or what is an acceptable standard, can only be determined subjectively by the school in terms of its own objectives and emphases and in terms of what may reasonably be expected of its students" (1971:534).

The two sets of achievement tests previously discussed are by no means the only ones available. For a source of available achievement tests in occupational education, see Boyd and Shimberg (1971).

According to Glaser and Nitko,

> If one assumes that measures of entering behavior can be obtained and that instructional treatments are available, then at the present state of knowledge, empirical work must take place to determine those measures most efficient for assigning individuals to classes of instructional alternatives (1971:644).

Aptitude tests have been used to predict the likelihood of success for a given student in a given instructional task or course of instruction. The basic assumption is that not all students will fare equally well in a given course as a consequence of differential entering behaviors.

The 1962 edition of the Differential Aptitude Tests consists of two forms, L and M. The test battery consists of the following tests:

1) Verbal Reasoning,
2) Numerical Ability,
3) Abstract Reasoning,
4) Clerical Speed and Accuracy,
5) Mechanical Reasoning,
6) Space Relations, and
7) Language Usage.

Quereshi, et al. assert that the Differential Aptitude Test (DAT) has not proven its ability to differentiate. They feel that the proper evidence should present "...(a) the appropriate combination of scores which sets one occupational group apart from another, and (b) the contribution that a particular test makes to the discriminant function identifying people in a particular occupation" (1972:1051). In conclusion, Quereshi, et al. state "...the DAT, if certain steps are taken, has better chances of attaining an acceptable level of differential efficiency than any other comparative battery" (1972:1052).

The General Aptitude Test Battery, B-1001 edition, consists of eight paper-and-pencil test plus four performance tests. They are:

1) Tool Matching,

2) Name Comparison,

3) Computation,

4) Three Dimensional Space,

5) Arithmetic Reasoning,

6) Vocabulary,

7) Form Matching,

8) Mark Making,

9) Pegboard Place,

10) Pegboard Turn,

11) Finger Dexterity Assembly, and

12) Finger Dexterity Disassembly.

Weiss, et al. say the General Aptitude Test Battery (GATB) "...leaves much to be desired if it is to be adequately used for vocational guidance" (1972:1058). They criticize the "pass-fail" qualifying score used for determining an individual's "Occupational Aptitude Pattern," and argue for at least a table of "hit rates" for each validity coefficient

reported. They criticize the tests as being too speeded and that the times are purely for administrative convenience. In their opinion, the test is somewhat outdated and the producers should "...immediately embark on a program to greatly expand the number and kinds of abilities measured by the GATB if the test battery is to have more than minor utility in the next decade" (Weiss, et al., 1972:1060).

In the opinion of this author, the same caveat about interpreting norms applies to these test batteries as it did to the achievement tests.

These are not the only aptitude measures available to the evaluator but are probably the most commonly used ones.

A test that may warrant further study by local educators is the Armed Services Vocational Aptitude Battery (ASVAB). This single test battery replaces the several tests used by different branches of the service in the past.

Test administration "...imposes no obligation on the part of school officials or students and involves no cost to local governments" (U.S. Department of Defense, 1968:2). Testing and scoring are taken care of by the representatives of the military services. The test battery requires about 2 1/2 hours for administration.

The ASVAB consists of nine component tests, each of which is of the pencil-and-paper variety. The tests generally consist of 25 items, with each item containing four alternative responses. The nine components tests are:

1) Coding Speed Test,

2) Word Knowledge,

3) Arithmetic Reasoning,

4) Tool Knowledge,

5) Space Perception,

6) Mechanical Comprehension,

7) Shop Information,

8) Automotive Information, and

9) Electronics Information.

The High School Counselors Manual provides information about how the component tests relate to armed services occupational groups and related civilian occupational fields from the Dictionary of Occupational Titles.

Certainly, questions exist which need additional study before decisions can be made about the usefulness of the ASVAB to the local district, but further study is currently being conducted.

## Interest Tests

According to Schwarz, "Because even the most accurate aptitude tests account for less than half of the variance in individual attainment, there has been a continuing effort to devise supplementary instruments to predict portions of the variance remaining." This has brought forth increased emphasis on the interest test which, he says, "...attempts to measure the relative reinforcement an individual derives from various types of activities, assuming, on the basis of considerable evidence, that satisfaction can be an important determinant of attainment" (Schwarz, 1971:317).

Interest tests are usually composed of a large number of items appearing in one of two formats: (1) Likes, Indifferent, or Dislikes (LID); and (2) triad. In the LID format, the individual is asked to respond to an item by marking "likes, indifferent, or dislikes." In the triad format, the individual is given three statements and asked to choose the one liked best and the one liked least. The LID format is used in the Strong Vocational Interest Blank (SVIB), while the triad format is used in both the Kuder Preference Record--Occupational, and the Minnesota Vocational Interest Inventory. Current research seems to give a slight advantage to the triad format (Berdie and Campbell, 1968).

The Strong Vocational Interest Blank for Men is designed for use with males aged 16 and over. It contains 84 scoring scales with 22 basic interests, 54 occupational, and eight nonoccupational, along with six administrative indices. The test, originally developed in 1927, has undergone many revisions,

the latest being in 1969. It is a paper-and-pencil instrument which takes about 40 minutes to administer. According to Krauskopf, the test is constructed according to the rationale that

> ...it is possible to differentiate men in a given occupation from men-in-general by asking questions about their likes and dislikes--and, further, that a person who likes and dislikes the same things as successful people in that occupation will be more likely to enter the occupation and be more likely to succeed in it (1972:1465).

There is a companion, but more occupationally limiting, test for women.

The Kuder Occupational Interest Survey (KOIS) contains the same items as the Kuder Preference Record--Occupational, but it is scored differently. The test is applicable in grades 11-16 and for adults. It has 106 scales for men and 84 scales for women. The test consists of 100 items written in the triad format. The latest revision is form DD. The reading level is approximately the sixth grade. An individual's scores are related to those of people in various occupations and fields of study (Dolliver, 1972).

These are not the only vocational interest tests, but they are the ones most commonly used and, apparently, from reading the reviews in The Seventh Mental Measurements Yearbook, the best of the lot.

For a comparison of the two tests, the following are excerpts from reviews of the Kuder Occupational Interest Survey:

> Inevitably, a comparison must be made between the Kuder DD and the SVIB. The DD has these advantages: (a) scoring of college major interests, (b) having a broader range of occupations (more technical and trade level occupations), (c) using the same test for males and females, (d) providing scores for female test takers on selected male occupational and college major scales, and (e) having norm groups which were more recently tested. But because the SVIB has accumulated more supporting reliability and validity data, the SVIB remains the better test, in this reviewer's opinion (Dolliver, 1972:1429).

> Thus, until additional validity and reliability data are accumulated for the KOIS, practitioners will probably be assuming less risk by using the more soundly researched SVIB (Walsh, 1972:1431).

Summary

With some glaring omissions, the number of achievement tests designed to measure the attainment of students in various vocational courses is growing. However, as such courses in these areas (e.g., plastics) become more widely offered, tests will be developed for them.

Aptitude measures seem lacking in both their predictive ability and their usefulness as guidance tools. However, the situation is improving: the existing ones are constantly being upgraded and new ones developed.

Both the aptitude and achievement tests should be carefully studied before use at the local level to ensure that: (1) the objectives the test is purporting to measure are applicable to the local situation, and (2) the referent group used to norm the test is similar enough to the local population to make comparisons meaningful.

In the area of measuring vocational interests, judging from the reviews in The Seventh Mental Measurement Yearbook, the Strong Vocational Interest Blank seems to be preferred. Again, though, the evaluator should judge the merits of any assessment instrument in the light of local needs.

## LOCALLY DEVELOPED MEASURES

What happens if the evaluator examines available tests and finds that they are not appropriate for the local situation? One possibility is to consider developing the measures locally. This section of the paper will discuss how to go about it.

As Thorndike says:

> Central to any test development enterprise, and in fact to any educational enterprise, is a clear, explicit statement of the objectives that the program is designed to achieve and that, in consequence, the test should be expected to assess (1971:8).

The statement of goals and objectives in education is certainly not new to education; however, perhaps the clarity and preciseness being demanded today is.

> Clarity of objectives is important in the educational enterprise at a number of levels and in a number of contexts. Clear objectives give direction to the curriculum maker in choosing from the wide range of content and from the multiplicity of media for presenting that content. They give direction to the teacher in planning a unit of instruction. They provide focus for the evaluator and test maker whose concern is to determine the extent to which the purposes of an educational program are being achieved (Krathwohl, 1971:17).

Objectives have been variously defined. For this discussion, objectives will be limited to behavioral objectives or performance goals and the terms will be treated as synonyms. A performance goal has been defined as "...An educational objective that clearly states measurable and observable performance (with tolerances) that identifies for the student and teacher the conditions under which the events or steps involved in learning will take place..." (Byers, 1971:3). Walbesser (1970) lists the six components of a behavioral objective as:

1) Who is to exhibit the behavior?

2) What observable performance (action) is the learner expected to observe?

3) What conditions, objects, and information is given?

4) Who or what initiates the learner's performance?

5) What responses are acceptable?

6) What special restrictions are there on the acceptable response?

It is from these specific objectives that the evaluator will determine the content from which to develop the items for the measurement instrument. In the opinion of this writer, the Walbesser manual is quite useable as an instructional guide for developing behavioral objectives. Writing Performance Goals: Strategy and Prototypes lists the following as advantages of performance goals.

1) Properly expressed goals permit any student to select the material or instructional content he needs on the basis of his present knowledge and skill for learning each new topic....

2) Statements of performance goals also permit educational objectives, tests, or examinations to be precisely correlated....

3) Performance goals permit the development of well-defined, short learning sequences and curricula, and identifiable conditions of learnings, as well as clearly defined relevant goals, achievement opportunity, and unambiguous evaluation stated in performance terms....

4) Clear performance goals permit the student to learn something he does not know. He is not forced to repeat that which he already knows... (Byers, 1971:4).

The overall procedure for preparing performance goals is presented in Figure 1.

Writing Performance Goals: Strategy and Prototypes (Byers, 1971) provides a step-by-step analysis of each of the steps listed in Figure 1. The context for their analyses is always the field of vocational and technical education. The presentation contains some detailed examples of prototypes of performance goals in agricultural education, business and distributive education, health education, technical education and trade and industrial education.

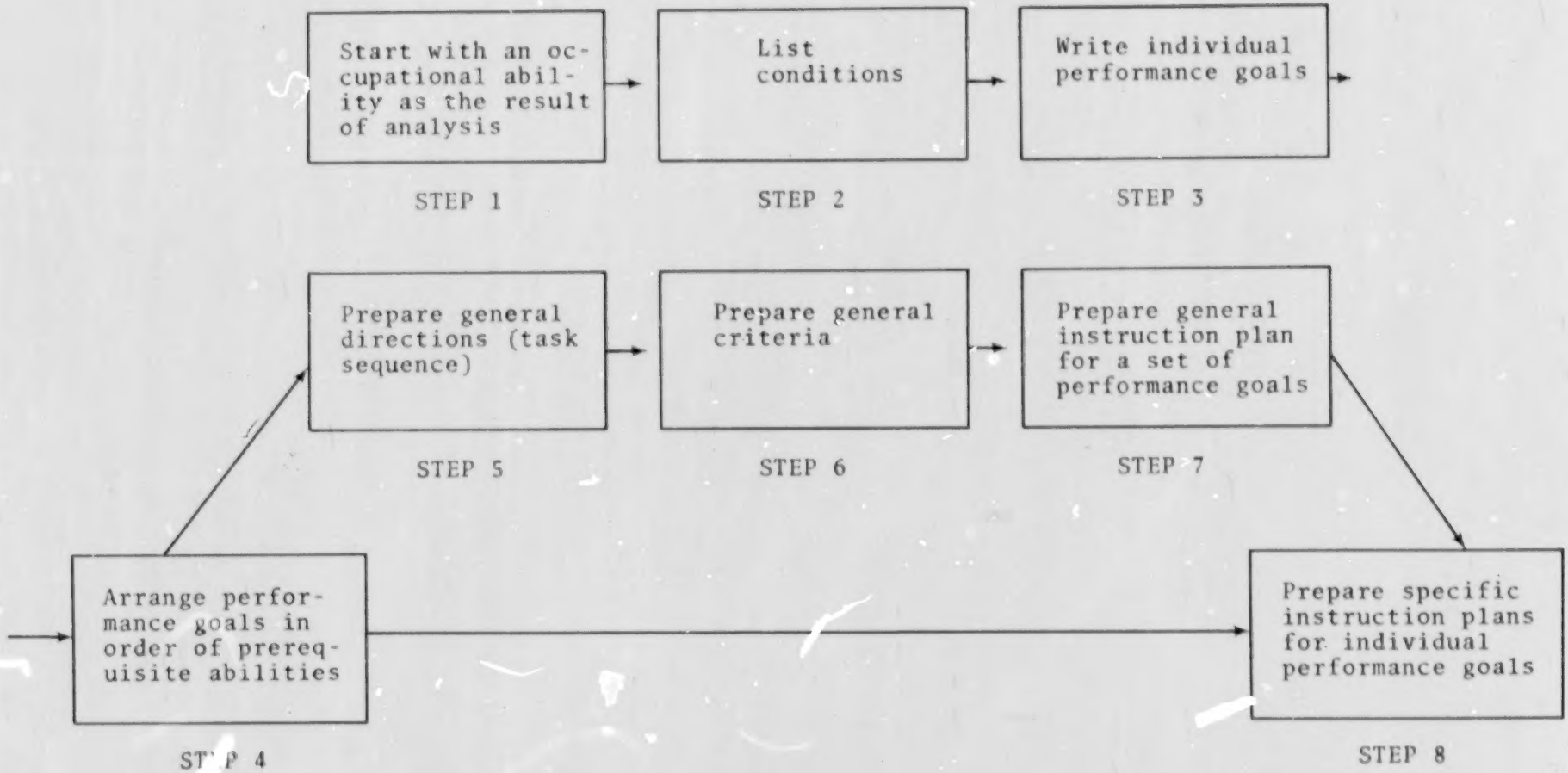
Start with an occupational ability as the result of analysis
List conditions
Write individual performance goals
STEP 1
STEP 2
STEP 3
Prepare general directions (task sequence)
Prepare general criteria
Prepare general instruction plan for a set of performance goals
STEP 5
STEP 6
STEP 7
Arrange performance goals in order of prerequisite abilities
Prepare specific instruction plans for individual performance goals
STEP 4
STEP 8



Figure 1. Overall Procedure for Preparing Performance Goals (Byers, 1971:5)

Once the objectives have been designed in a "performance goal" format, the evaluator must develop a test to assess the expected outcomes. Developing a test requires a systematic process of developing, analyzing and redeveloping. Krathwohl and Payne (1971:20) recommend a process which consists of the following 12 steps.

1) Specify the ultimate goals of the education process.

2) Derive from these the goals of the portion of the system under study.

3) Specify these goals in terms of expected student behavior. If relevant, specify the acceptable level of successful learning.

4) Determine the relative emphasis or importance of various objectives, their content and their behaviors.

5) Select or develop appropriate situations that will elicit the desired behavior in the appropriate context or environment, assuming the student has learned it.

6) Assemble a sample of such situations so that together they best represent the emphasis on content and behavior previously determined.

7) Provide for the recording of responses in a form that will facilitate scoring but that does not change the nature of the behavior elicited so that it is no longer a true sample or an accurate index of the behavior desired.

8) Establish scoring criteria and guides to provide objective and unbiased judgments.

9) Try out the instrument in preliminary form.

10) Revise the sample of situations on the basis of tryout information.

11) Analyze reliability, validity, and score distribution in accordance with purposes of score use.

12) Develop test norms and a manual, and reproduce and distribute test.

In the opinion of this author, each of these steps is vitally important, with the exception of the last one, to the successful development of a locally devised test. It is an effort that requires a great deal of time and resources but any attempts at shortcutting the systematic process will only reward the evaluator with an inferior, if not invalid, instrument.

The measurement instrument developed to assess the expected outcomes will usually be of the paper-and-pencil variety or a hands-on performance test.

In the paper-and-pencil test, the item format can vary from true/false, to multiple-choice, to an essay variety. "The multiple-choice form is by far the most popular one in current use" (Wesman, 1971:94). This form consists of an introductory question or an incomplete statement followed by two or more possible responses. The introductory portion is called the "stem" and the incorrect choices among the possible responses are called the "distractors." The number of distractors to include in an item is dependent upon the amount of time anticipated for the test, the nature of the item, the age group for which the test is intended, and other factors. "Item writers should try conscientiously to produce three or four distractors for multiple-choice items" (Wesman, 1971:102).

Wesman offers the following general suggestions for item writers.

1) Express the item as clearly as possible....

2) Wherever possible, choose words that have precise meanings....

3) Avoid complex or awkward word arrangements....

4) Include all qualifications needed to provide a reasonable basis for response selection....

5) Avoid the inclusion of nonfunctional words....

6) Avoid unessential specificity in the stem or the response....

7) Be as accurate as possible in all parts of an item....

8) Adapt the level of difficulty of the item to the group...for which it is intended....

9) Avoid irrelevant clues to the correct response....

10) Avoid stereotyped phraseology in the stem or the correct response....

11) Avoid irrelevant sources of difficulty....

12) Expose items to expert editorial scrutiny... (Wesman, 1971:102-111).

For a detailed discussion of each of the above suggestions along with examples, the reader is encouraged to refer to the original work cited.

A great deal of the vocational and technical education content lends itself well to, indeed, possibly dictates, that performance tests be constructed to assess attainments rather than the paper-and-pencil test. A performance test has been defined as "...one in which some criterion situation is simulated to a much greater degree than is represented by the usual paper-and-pencil test" (Fitzpatrick and Morrison, 1971: 238). In the vocational and technical education field, a criterion situation might be anything from operating a piece of machinery to typing a letter.

> Today it is generally conceded that written tests of trade knowledge are not a very dependable way to evaluate shop performance and that without some type of direct or indirect measure of actual performance it is unlikely that we can make an accurate assessment of an individual's trade competency (Boyd and Shimberg, 1971b:2).

Thus the primary value of a performance test is its ability to assess a skill in a situation approximating the real world. The disadvantages of performance testing are that: (1) they usually must be given to one student at a time, with limited equipment available; (2) it takes a great deal of time to test a group of students; (3) the real job situation

is often impractical to reproduce; and (4) students can easily pass along vital information to other students about the test (Boyd and Shimberg, 1971b).

Fitzpatrick and Morrison (1971) describe the process of developing a performance test as being similar to that used in developing other tests. They assert that specific information is needed about the job to be simulated (including cues for each action), environmental and social conditions under which the simulation is to occur, and information specific to the appearance and functioning of the equipment to be used. The task to be performed must be described in detail, which requires a thorough task analysis. The task analysis should provide the following information:

1) the initiating condition,

2) the action or response, and

3) the terminating condition.

From the task analysis, the evaluator has a thorough description of the job to be simulated. From this information, the performance goal is developed which is then used to develop the specific performance measure. It may be necessary to sample the behaviors within a given job because of practical time limitations. For this reason, it is necessary to have the various activities rated as to how critical they are. It may also be necessary to give a student a job completed up to a certain point and ask him to complete it. Once the test content has been decided on, a set of explicit instructions should be developed, including how the student's performance will be scored. Equipment and materials needed for the testing procedures should be carefully detailed.

A basic question arises relative to performance testing: is the process which a student uses most important, or is the finished product most important? Or, are they both important? This is a question that is situation specific. Part of the answer depends on the measurability of both the process and product, and part depends upon the importance placed on the two alternatives by experts.

The test administrator plays an active role in performance testing. He may even be called upon to act out a role in the simulation. It is his duty to see that conditions are the same

for all who take the test. If it is necessary to have an observer for the performance test, then he must be told what to observe, how to classify his observations, and how to record them. Scoring of the test must be determined by assigning the appropriate number of points to each particular activity within the performance test. These points can be determined by any number of methods, ranging from subjective judgments to a sophisticated statistical method (Fitzpatrick and Morrison, 1971).

Fitzpatrick and Morrison claim that:

> The potential value of the performance test lies in its closer approach to reality--its greater relevance in determining the degree to which the examinee can actually perform the tasks of the criterion job or other situation (1971:268).

They go on to say that the value is not obtained without a considerable expenditure of resources and some loss in test reliability. They suggest that:

> If an adequately relevant and otherwise suitable paper-and-pencil test is available or can readily be developed, there is no point in using or developing a performance test. However, the ready availability of paper-and-pencil tests has often blinded us to considerations of relevance.... Relevance is the primary consideration, and good measurement is only a means to the end of appropriate evaluation (1971:268).

Authors generally agree that one type of behavioral objective, the affective objective, is the most difficult to teach and to evaluate. Banks has said:

> ...of the three behavioral domains--cognitive, psychomotor, and affective--the affective is the most perplexing of all.
>
> Yet the attainment of acceptable, specified affective behaviors by students is of concern not only to vocational-technical educators but to those responsible for funding our educational institutions--and certainly to employers (1973: 36).

Tuckman accuses schools of having concentrated on occupational exploration and the dissemination of career information to the detriment of the development of proper student attitudes and motivation (Tuckman, 1973). Different reasons are given as to why schools are not stressing the affective domain. Banks says it is because we are not able as educators to write affective performance objectives because we do not know what the employer requires (Banks, 1973). Another reason given is that there exists a fear of brainwashing to indoctrination (Bloom, et al., 1971). Tuckman (1973) proposes that career development be expanded to include the affective dimension. He wants teachers to be assisted through either course work or curriculum guides so that they can effectively teach the development of affective behaviors.

The problems with trying to define in a performance goal manner the expected behaviors in the affective domain illuminate the problems of trying to measure those objectives. Banks (1973) offers examples of some performance goals and some methods for evaluating them. The evaluation consists of the use of structured scales for recording subjective assessments. Bloom, et al. (1971) make the distinction between evaluating the affective goals of curriculum and the affective behavior of individuals. The basic tenet of the distinction is that to evaluate the curriculum goals the individual anonymity can be assured, whereas with evaluation of the individual behavior it cannot be. He offers examples of different techniques for measuring affective objectives. He concludes:

> Models of well-defined affective objectives and a variety of techniques to evaluate them are available to the teacher or school system willing to accept the obligation to assess previously neglected but important affective curriculum components (Bloom, et al., 1971:244).

Using objectives from the Instructional Objectives Exchange, Giguere and Baker (1971) have developed criterion-referenced measures for the assessment of attitudes toward school and self-concept. The school attitude measures consisted of six dimensions:

1) teachers,

2) school subjects,

3) learning,

4) school social structure and climate,

5) peer, and

6) general.

"Seventeen objectives were identified in the self-concept realm, six at the primary level, six at the intermediate level, and five at the secondary level" (Giguere and Baker, 1971:3). The measures developed were intended for assessing group attitudes toward school and self-concept rather than individual attitudes. After a trial on a sample of students, the measures were revised and then field tested. The analysis of field test results showed that the instruments were satisfactory for making group decisions. For a technical analysis of the measures, the reader is referred to Giguere and Baker (1971).

Many different methods of attaining measures of student attitudes are being developed and field tested. The methods vary widely in design and intent. The evaluator who is faced with the challenge of measuring student attitudes should be rewarded by a careful review of what has been, and is being done, in the field. As an example of the variety of approaches being taken, the author cites a few studies which have used different methods for measuring student attitudes.

Murray (1971) has used the Thurstone approach for measuring attitudes which involves developing a series of statements about a topic and asking students to respond to these statements.

Estes (1972) developed a series of statements about reading and asked students to respond to the statements using a Likert scale. He presented the following 14 criteria for item writing.

1) Avoid statements referring to past rather than present.

2) Avoid factual statements.

3) Avoid ambiguity.

4) Avoid statements irrelevant to the psychological object under consideration.

5) Avoid statements likely to be endorsed by almost anyone or no one.

6) Select statements believed to cover entire range of affective interest.

7) Use simple, clear, direct language.

8) Make statements short--20 words.

9) Each statement should have only one complete thought.

10) Avoid all, always, none, and never--they are ambiguous.

11) Words like only, just, or merely are to be avoided.

12) Use simple sentences.

13) Avoid use of words perhaps incomprehensible to group.

14) Avoid use of double negatives (Estes, 1972: 5, 6).

Hamersma (1971) discusses the Guttman fact design and analysis technique for developing attitude scales. This technique allows the evaluator to develop items in a systematic, a priori design.

Harvill (1971) examines the effectiveness of five different methods of measuring attitudes of young children. The study compares two ipsative measures, the Picture method and a triad forced choice method. Three other response methods compared were the Millimeter, Box and Semantic Differential. Harvill concludes that:

1) Ipsative attitude measures should be used with great caution.

2) Teacher nominations...are...not very valid as a criterion measure....

3) The two most promising...methods are...the Millimeter and Semantic Differential... (Harvill, 1971).

In summary, the first step in developing a measure of student achievement or attitude is the careful statement of the objectives. The format for behavioral objectives or performance goals proposed by the authors referred to in this document certainly provides a model for the local evaluator to consider. Development of performance goals is a time consuming activity. However, Walbesser and Eisenberg (1972), in a review of research on the effectiveness of using behavioral objectives, seem to give a slight advantage toward using behavioral objectives. There is a need for continued research on the subject. The literature is heavily weighted toward the development and use of paper-and-pencil tests for the assessment of student achievement. This is unfortunate, for in the opinion of this author, performance testing is more relevant for many of the activities in the field of vocational and technical education. Many of the problems encountered in trying to assess the affective domain of behavior can be traced to a lack of explicit objectives. However, efforts are being made in this field and some results, at least in measuring group attitudes, have been reported.

## EPILOGUE

The central assumption of this paper has been that the evaluation of students in vocational and technical education programs is directly related to, indeed, inseparable from problems inherent in selecting or adopting proper assessment instruments. As pointed out before, Thorndike believes that:

> Central to any test development enterprise and in fact to any educational enterprise, is a clear, explicit statement of the objectives that the program is designed to achieve and that in consequence, the test should be expected to assess (1971:8).

Thus, the first problem faced by the local educator who is attempting to evaluate students in vocational and technical education programs is to obtain a measurable set of

objectives. These objectives can then be matched with those for which the test is written.

Another problem is the determination of the comparability of the sample of students used in norming a selected test to the local student population. If there are serious doubts as to the comparability of the groups, then local norms may need to be developed.

The leading tests of achievement, aptitude and interests of vocational and technical education students have been reviewed and reported. It is apparent from this review that good measures do exist and warrant careful consideration from the local evaluator.

This author encourages local evaluators to conduct a careful survey of the literature before attempting to develop any local measures. A great deal of work has been done; it would certainly be unwise to waste valuable resources in "reinventing the wheel." A review of the Education Resources Information Center (ERIC) system is a logical place to begin.

# BIBLIOGRAPHY [1]


Angoff, William H. "Scales, Norms, and Equivalent Scores." Educational Measurement. Second Edition. Edited by Robert L. Thorndike. Washington, D.C.: American Council on Education. 1971.

Baldwin, Thomas. The Development of Achievement Measures for Trade and Technical Education. Final Report. Urbana: University of Illinois. 1970. 109 pp. ED 044 526 MF $0.65 HC $6.58.

Banks, Jane M. "Measuring the Immeasurables." American Vocational Journal. Vol. 48, No. 4 (April, 1973), 36-37.

Berdie, Ralph F., and Campbell, David P. "Measurement of Interest." Handbook of Measurement and Assessment in Behavioral Sciences. Edited by D. K. Whitla. Reading, MA: Addison-Wesley Publishing Company. 1968.

Bloom, Benjamin S., et al. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill. 1971.

Boyd, Joseph L., Jr., and Shimberg, Benjamin. Directory of Achievement Tests For Occupational Education. Princeton, NJ: Educational Testing Service. 1971a.

_____. Handbook of Performance Testing: A Practical Guide for Test Makers. Princeton, NJ: Educational Testing Service. 1971b.


---

[1] Bibliographical entries followed by an ED number are generally available in hard copy or microfiche through the Educational Resources Information Center (ERIC). This availability is indicated by the abbreviations MF for microfiche and HC for hard copy. Order from ERIC Document Reproduction Service (EDRS), P.O. Drawer O, Bethesda, Maryland 20014. Payment must accompany orders totaling less than $10.00.

Buros, Oscar K., ed. The Seventh Mental Measurements Yearbook. Volumes I, II. Highland Park: The Gryphon Press. 1972.

Byers, Edward E., Director. Writing Performance Goals: Strategy and Prototypes. A Manual for Vocational and Technical Educators. New York: McGraw-Hill Book Company, Gregg Division; and Columbus: The Center for Vocational and Technical Education, The Ohio State University. 1971. 107 pp. ED 061 413 MF $0.65 HC $6.58.

Carpenter, P., and Rapp, M. Testing in Innovative Programs. Santa Monica: Rand Corporation. 1972. 10 pp. ED 068 576 MF $0.65 HC $3.29.

Dolliver, Robert H. "Tests and Reviews: Vocations-Interests." The Seventh Mental Measurements Yearbook. Vol. II. Edited by Oscar K. Buros. Highland Park, NJ: The Gryphon Press. 1972.

Estes, Thomas H. "Assessing Attitudes Toward Reading." Paper presented at the meeting of International Reading Association, Detroit, Michigan, May 11, 1972. May, 1972. 10 pp. ED 063 077 MF $0.65 HC $3.29.

Fitzpatrick, Robert, and Morrison, Edward J. "Performance and Product Evaluation." Educational Measurement. Second Edition. Edited by Robert L. Thorndike. Washington, D.C.: American Council on Education. 1971.

Giguere, Constance L., and Baker, Eva L. The Development of Criterion-Referenced Tests of Self-Concept and Attitude Toward School. Los Angeles: University of California. 1971. 14 pp. ED 054 192 MF $0.65 HC $3.29.

Glaser, Robert, and Nitko, Anthony J. "Measurement in Learning and Instruction." Educational Measurement. Second Edition. Edited by Robert L. Thorndike. Washington, D.C.: American Council on Education. 1971.

Hamersma, Richard J. Guttman Facet Design and Analysis: A Technique for Attitude Scale Construction. 1971. 13 pp. ED 058 286 MF $0.65 HC $3.29.

Harvill, Leo M. Evaluation of Several Methods for Measuring Young Children's Educational Attitudes. Vermillion: School of Education, South Dakota University. 1971. 38 pp. ED 056 059 MF $0.65 HC $3.29.

Krathwohl, David R., and Payne, David A. "Defining and Assessing Educational Objectives." Educational Measurement. Second Edition. Edited by Robert L. Thorndike. Washington, D.C.: American Council of Education. 1971.

Krauskopf, Charles J. "Tests and Reviews: Vocations-Interests." The Seventh Mental Measurements Yearbook. Vol. II. Edited by Oscar K. Buros. Highland Park: The Gryphon Press. 1972.

Murray, Norman B. "Construction of a Thurstone Attitude Test." Paper presented at the Annual Convention of the California Educational Research Association, San Diego, California, April, 1971. April 30, 1971. 16 pp. ED 058 306 MF $0.65 HC $3.29.

Ohio Trade and Industrial Education Services. Using the Results of the Trade and Industrial Education Achievement Test Program. Columbus: Instructional Materials Laboratory, The Ohio State University. 1972.

Quereshi, M. Y., et al. "Multi-Aptitude Batteries." The Seventh Mental Measurements Yearbook. Vol. II. Edited by Oscar K. Buros. Highland Park, NJ: The Gryphon Press. 1972.

Schwarz, Paul A. "Prediction Instruments for Educational Outcomes." Educational Measurement. Second Edition. Edited by Robert L. Thorndike. Washington, D.C.: American Council on Education. 1971.

Seibel, Dean W. "Measurement of Aptitude and Achievement." Handbook of Measurement and Assessment in Behavioral Sciences. Edited by Dean K. Whitla. Reading, MA: Addison-Wesley Publishing Company. 1968.

Thorndike, Robert L. "Educational Measurement for the Seventies." Educational Measurement. Second Edition. Edited by Robert L. Thorndike. Washington, D.C.: American Council on Education. 1971.

Tuckman, Bruce W. "Getting at the Attitudes Problem." American Vocational Journal. Vol. 48, No. 1 (January, 1973), 47-48.

U.S. Department of Defense. Armed Services Vocational Aptitude Battery: School Counselors Manual. Washington, D.C.: Superintendent of Documents, 1968.

Walbesser, Henry H. Constructing Behavioral Objectives. Second Edition. College Park, MD: Bureau of Educational Research and Field Services, University of Maryland. 1970.

_____, and Eisenberg, Theodore A. A Review of Research on Behavioral Objectives and Learning Hierarchies. Columbus: ERIC Information Analysis Center for Science, Mathematics and Environmental Education. 1972.

Walsh, W. Bruce. "Tests and Reviews: Vocations-Interests." The Seventh Mental Measurements Yearbook. Vol. II. Edited by Oscar K. Buros. Highland Park, NJ: The Gryphon Press. 1972.

Weiss, David J., et al. "Multi-Aptitude Batteries." The Seventh Mental Measurements Yearbook. Vol. II. Edited by Oscar K. Buros. Highland Park, NJ: The Gryphon Press. 1972.

Wesman, A. G. "Writing the Test Item." Educational Measurement. Second Edition. Edited by Robert L. Thorndike. Washington, D.C.: American Council on Education. 1971.



U.S. GOVERNMENT PRINTING OFFICE: 1974 - 737-914/5906